Basic Concept

Annotate Corporate Social Responsibility (CSR) statements to evaluate their authenticity, helping consumers and investors distinguish genuine sustainability efforts from greenwashing.

Label Categories

L1: Genuine Impact-Driven
L2: Mixed (Real Impact but PR-heavy)
L3: Pure Greenwashing

Source of Data

Publicly available, non-proprietary, and non-copyrighted

- corporate social responsibility reports
- public corporate press releases
- sustainability blogs
- company websites

Model Accuracy

Best Model: *BERT Fine-Tuning*

Test Accuracy: 99%

95% Confidence Interval: (0.9703, 1.0097)

*Theoretically, accuracy should not be > 100%, this is just an approximate statistical error

	Model	Dev Accuracy	Test Accuracy	95% Confidence Interval
0	Majority Classifier	0.52	0.52	(0.4209, 0.6191)
1	Logistic Regression (TF-IDF)	0.92	0.92	(0.8662, 0.9738)
2	Ordinal Regression (TF-IDF)	0.96	0.95	(0.9068, 0.9932)
3	BERT Fine-Tuning	0.99	0.99	(0.9703, 1.0097)

Interesting Investigation & Analysis

- Importance of language features
  - Sentences with "verified", "results, "third-party verification" and "specific numbers" are almost always correctly classified as Genuine.
  - Excessive use of abstract words such as "journey", "aspirations", "planet" are easily classified as Greenwashing.

- Dataset Imbalance Effects: Greenwashing (Label 3) is the least frequent class, causing imbalance
- Model Generalization: BERT generalizes well even under subtle language shifts